

# Sensitivity to the prototype in children with high-functioning autism spectrum disorder: An example of Bayesian cognitive psychometrics

Wouter Voorspoels<sup>a</sup>, Isa Rutten<sup>a</sup>, Annelies Bartlema<sup>a</sup>, Francis Tuerlinckx<sup>a</sup>,  
Wolf Vanpaemel<sup>a,\*</sup>

<sup>a</sup>*University of Leuven, Belgium*

---

## Abstract

We present a case study of hierarchical Bayesian explanatory cognitive psychometrics, examining information processing characteristics of individuals with high-functioning Autism Spectrum Disorder (HFASD). On the basis of previously published data, we compare the classification behavior of a group of children with HFASD with that of typically developing (TD) controls using a computational model of categorization. The parameters in the model reflect characteristics of information processing that are theoretically related to HFASD. Because we expect individual differences in the model's parameters, as well as differences between HFASD and TD children, we use a hierarchical explanatory approach. A first analysis suggests that children with HFASD are less sensitive to the prototype. A second analysis, involving

---

\*Corresponding author. Address: Faculty of Psychology and Educational Sciences, Tiensestraat 102, bus 3713, 3000 Leuven, Belgium. Email: [wolf.vanpaemel@kuleuven.be](mailto:wolf.vanpaemel@kuleuven.be), +3216326256, Fax: +3216325993

*Email addresses:* [wouter.voorspoels@kuleuven.be](mailto:wouter.voorspoels@kuleuven.be) (Wouter Voorspoels), [isa.rutten@kuleuven.be](mailto:isa.rutten@kuleuven.be) (Isa Rutten), [annelies.bartlema@ppw.kuleuven.be](mailto:annelies.bartlema@ppw.kuleuven.be) (Annelies Bartlema), [francis.tuerlinckx@kuleuven.be](mailto:francis.tuerlinckx@kuleuven.be) (Francis Tuerlinckx), [wolf.vanpaemel@kuleuven.be](mailto:wolf.vanpaemel@kuleuven.be) (Wolf Vanpaemel)

a mixture component, reveals that the computational model is not appropriate for a subgroup of participants, which implies parameter estimates are not informative for these children. Focusing only on the children for whom the prototype model is appropriate, no meaningful difference in sensitivity between HFASD and TD children is inferred.

*Keywords:*

---

## **1. Introduction**

Autism Spectrum Disorder (ASD) is characterized by difficulties with reciprocal social interaction, abnormalities in communication, non-functional, restricted interests, and repetitive and stereotyped behaviors (American Psychiatric Association, 2013). While the socio-communicational symptoms are salient in everyday functioning, it has been suggested that cognitive characteristics play a key role in the aetiology of the phenotype of ASD (?). Of particular interest are the difficulties that individuals with ASD experience when transferring knowledge from familiar to novel situations. Several leading theories (Happé & Frith, 2006; Motttron et al., 2006; Plaisted, 2001; ?) propose that individuals with ASD have enhanced perceptual discrimination abilities and form hyperspecific representations that are extremely detailed and focused on differentiating objects. Hyperspecific representation are thought to impede abstraction of more general knowledge and generalization to novel contexts, events or objects (Bott et al., 2006; Happé & Frith, 2006).

Church et al. (2010) examined whether high functioning (HF) children with ASD use abstract category information to the same extent when mak-



ing categorization as typically developing (TD) children. They presented a dot pattern categorization task to both TD and HFASD children and analyzed the response patterns with a computational model of categorization that assumes category knowledge is represented in the form of an abstracted prototype. On the basis of individual level estimates of the parameters in the computational model analysis, they compared categorical processing in HFASD children and matched TD controls. Church et al. found that HFASD children did not use abstracted category knowledge in the same way when classifying novel objects as TD children, in that they appeared less sensitive to differences between stimuli when generalizing.

In this paper, we take the lead from Church et al. (2010) to illustrate the application of hierarchical Bayesian explanatory cognitive psychometrics. In particular, we use an item response model that has strong theoretical foundations in cognitive science and was developed to understand the performance of people in categorization tasks, to assess individual differences in sensitivity to the prototype. Besides charting individual differences, we evaluate an explanatory component by examining differences at a group level between known groups, that is, typically developing children and children with Autism Spectrum Disorder. In our analyses, we rely on Bayesian inference for parameter estimation, we make use of hierarchical models to capture differences between the two a priori groups in the study (TD and HFASD children), and we add a mixture component to accommodate the possibility that some children relied on different strategies in the categorization task. We describe all steps of the Bayesian modelling approach, including parameter estimation and model comparison, evaluation and expansion.

The structure of this paper is as follows. First, we provide the necessary background on the modelling approach, followed by a description of the task, the data and the model used to assess differences in sensitivity. We then describe two analyses. In a first analysis, we recast the analysis of Church et al. (2010) in a hierarchical Bayesian framework. In a second analysis, we extend the modelling approach by implementing a mixture modeling component that isolates children for which the appropriateness of the model can be questioned. By not taking these children in account when making inferences about the parameters in the model, the inferences become more valid.

## **2. Hierarchical Bayesian explanatory cognitive psychometrics**

Our approach for studying individual differences relies on information processing models with a strong theoretical foundation in the cognitive sciences, considers explanatory covariates, and uses a hierarchical Bayesian approach to model evaluation and parameter inference. In this section, we discuss each of these aspects in more detail.

### *2.1. Cognitive psychometrics*

As a discipline, psychometrics is concerned with the measurement of psychologically interesting aspects of individuals. This involves inferring the position of individuals on latent variables on the basis of their responses to a set of questions, stimuli or items. From this definition, it is apparent that psychometrics is closely related to the study of individual differences. The early developments culminated in the classical test theory (see, e.g., the reference work of Lord & Novick, 1968). Although elegantly formulated and still widely used in practical settings, classical test theory rests on an untestable

mathematical decomposition (i.e., the “true score” theory) and is therefore becoming increasingly superseded by item response theory (IRT).

IRT is not a single theory, but rather a large family of general testable statistical models. All item response models incorporate some theory of what happens when an item is presented to a person. For many well-known models, the theoretical foundation is rather shallow: For example, in the Rasch model, both a person and an item are characterized by a single real number (generally referred to as ability and difficulty, respectively) and their difference determines the probability that the person answers the item correctly (i.e., a simple rule of domination). This corresponds to assuming a single dimension underlying the task, without any further elaboration on what it is that constitutes the dimension or which cognitive processes are involved when doing the task.

The theoretical foundation of an item response model can be considerably strengthened by considering the psychological processes underlying an individual’s responses (e.g., Tuerlinckx & De Boeck, 2005; van der Maas et al., 2011). Information processing models developed and used in the cognitive sciences are excellently suitable for this purpose. These models formally describe the processes underlying behavior in tasks involving thinking, remembering, perceiving, deciding, learning and so on. The application of cognitive process models as psychometric measurement models has been coined *cognitive psychometrics* (Batchelder, 1998; Riefer et al., 2002). The central idea in cognitive psychometrics is to apply sophisticated quantitative models of cognition to measure specific information processing characteristics, allowing precise and mathematical formulation of hypotheses, and sophisticated

quantitative tests.

## 2.2. *Explanatory psychometrics*

The majority of IRT applications are descriptive. Models are used to assess the individual differences, without explaining these differences in terms that go beyond the ones inscribed by the models' basic theoretical foundation. In line with the rise of multilevel (hierarchical) models (see Verbeke & Molenberghs, 2009), the framework of *explanatory psychometrics* has been developed. In explanatory item response models (e.g., De Boeck & Wilson, 2004), the individual differences are explained by relating them to external person variables, such as measures on traits or experimental conditions. The goal is not only to position individuals on latent variables that are meaningful parameters, but also to relate individual differences and group differences to potential covariates.

## 2.3. *Hierarchical mixture models*

Psychometric modelling has often involved building hierarchical models (Gelman & Hill, 2006; Kruschke & Vanpaemel, 2015; Lee, 2011; Rouder & Lu, 2005). A hierarchical model comprises two (or more) levels, each containing parameters that are inferred from observed data. At the lowest level, an item response model specifies how a person generates a response to an item on the basis of person-specific parameters. These parameters indicate the positions of the persons on the dimensions of interest. It is assumed that the person-specific parameters are drawn from a group distribution at a higher hierarchical level, defined by a number of group-specific parameters. The group distribution can be considered a population distribution for the

person-specific parameters and some of its features — such as the mean and variability — can be inferred from the data.

Hierarchical models enable researchers to consider population differences by investigating differences in the group-specific parameters (i.e., explaining the differences among persons using group membership as a person covariate). In contrast with more traditional approaches such as averaging the data within a group, or separately analyzing each participant, the hierarchical approach does not assume that all individuals within a group are identical, nor does it neglect similarities between individuals of the same population. All parameters are simultaneously estimated from the observations, capturing both the differences between individuals and the differences between populations, holding the middle ground between an approach that allows every individual to be unique and an approach that only considers population averages. Especially when only few data are available per participant, the strength of a hierarchical analysis is in partial pooling: The individuals' estimates within a group are used to infer the population characteristics, which are simultaneously used to reduce uncertainty in an individual's estimates.

Inter-individual differences are not restricted to different values of parameters in a computational model. While such quantitative differences are often the focus of interest, qualitative differences should also be considered (e.g., Bartlema et al., 2014). In the context of behavioral data from a cognitive task, it is plausible to assume that there are more than one way to go about the task. Not all participants rely on the same heuristics, strategies or cognitive processes when doing the task. Consequently, it is important to consider alternative strategies in the analysis, because parameter estimates

are only informative to the extent the model they figure in is appropriate. By adding mixture components, latent groups can be discovered, that is, clusters of individuals that appear to use a common strategy.

#### *2.4. Bayesian inference*

In hierarchical mixture models, inferences about parameters are typically done in a Bayesian fashion. Bayesian inference is generally acknowledged as a complete, coherent and intuitive way of relating data and theory, and its application in psychological research has taken flight over the last decade (Lee, 2008, in press). In a Bayesian view, knowledge about parameters is represented by probability distributions. In the present case, the parameters map onto relevant aspects of cognitive functioning – sensitivity – and the probability distribution reflects our knowledge and uncertainty about these parameters. Before observing any data, our knowledge of a parameter  $\theta$  is captured in a prior distribution  $p(\theta)$ , providing a quantification of the uncertainty about the values the parameter may take. Often, theory or logic bring expectations about the parameters, and the prior can be used to reflect this information (Lee & Vanpaemel, submitted; Vanpaemel & Lee, 2012). If there is no prior knowledge available about  $\theta$ , a broad and non-committal prior is chosen. When experimental data  $D$  become available, the prior distribution is updated, yielding the posterior distribution  $p(\theta|D)$  by applying Bayes' rule,  $p(\theta|D) = \frac{p(D|\theta)}{p(D)}p(\theta)$ . The posterior quantifies the current knowledge and uncertainty about  $\theta$  after having taken the data into account.

### 3. Autism and sensitivity to the prototype

Church et al. (2010) collected data in a dot pattern categorization task (e.g., Knowlton & Squire, 1993; Posner & Keele, 1968) from a group of HFASD children, and a group of TD children matched on age, gender and IQ. In a training phase, the children were presented with dot patterns and informed about which dot patterns were “cave ghosts”. In the test phase, they were asked to decide for a number of novel dot patterns whether they were cave ghosts. HFASD children were observed to endorse the category prototype significantly less and seemed to make less use of similarity to the prototype in their classification judgments. To examine this further, Church et al. fitted a prototype model (Nosofsky, 1987) to the observed response pattern of each participant separately. The prototype model assumes that during the training phase, abstract knowledge about the category is retained in the prototype, and that subsequent classification of novel stimuli relies on this abstracted prototype. Their model analyses revealed that the average sensitivity to the prototype, as quantified by a parameter in the prototype model, was substantially smaller for the HFASD children than for the TD children.

#### *3.1. Categorization Task*

##### *3.1.1. Procedure.*

The stimuli involved a prototype (P), 40 distortions (L) from the prototype varying in level of distortion, and 45 random stimuli (R). Each stimulus was a dot pattern (see Posner & Keele, 1968) in which the dots were connected by lines. There were five different levels of distortion, denoted as L2, L3, L4, L5 and L7, with the higher level having less resemblance to the

prototype. There were five L2, ten L3, five L4, ten L5, and ten L7 stimuli.

All participants completed a training phase comprising 30 trials, in which five L3, five L5, five L7, and 15 random, non-category member stimuli were presented. On each trial, participants were asked to indicate whether the presented stimulus was a “cave ghost” (that is, a category member), after which corrective feedback was given. During the test phase, five examples of the prototype and of each of the five distortion types were presented, as well as 30 random stimuli, totalling 60 trials. All stimuli were presented in a fixed order and none of the test stimuli were presented in the training phase. During the test phase, participants did not receive corrective feedback.

### *3.1.2. Data.*

The study by Church et al. (2010) involved 40 children ages 7 to 12 years, with 20 in each group (HFASD and TD). The two groups were matched on age, gender and IQ. The inclusion criteria for children in the HFASD group can be found in Church et al. (2010, p. 864). Figure 1 shows the average data pattern for both groups, as well as the observed pattern for each individual subject, in the form of the proportion of endorsement as a cave ghost for each stimulus type in the test phase. Each participant has been assigned an identification number that is consistent throughout the manuscript, 1 to 20 for TD children and 21 to 40 for children with HFASD.

The left panel of Figure 1 shows a clear general pattern, both for HFASD and TD children: Stimuli being more similar to the prototype are more readily endorsed as category members (cave ghosts). However, looking at the individual response patterns in the right panel, there are a number of children that do not seem to follow the general pattern (e.g., participants 3,



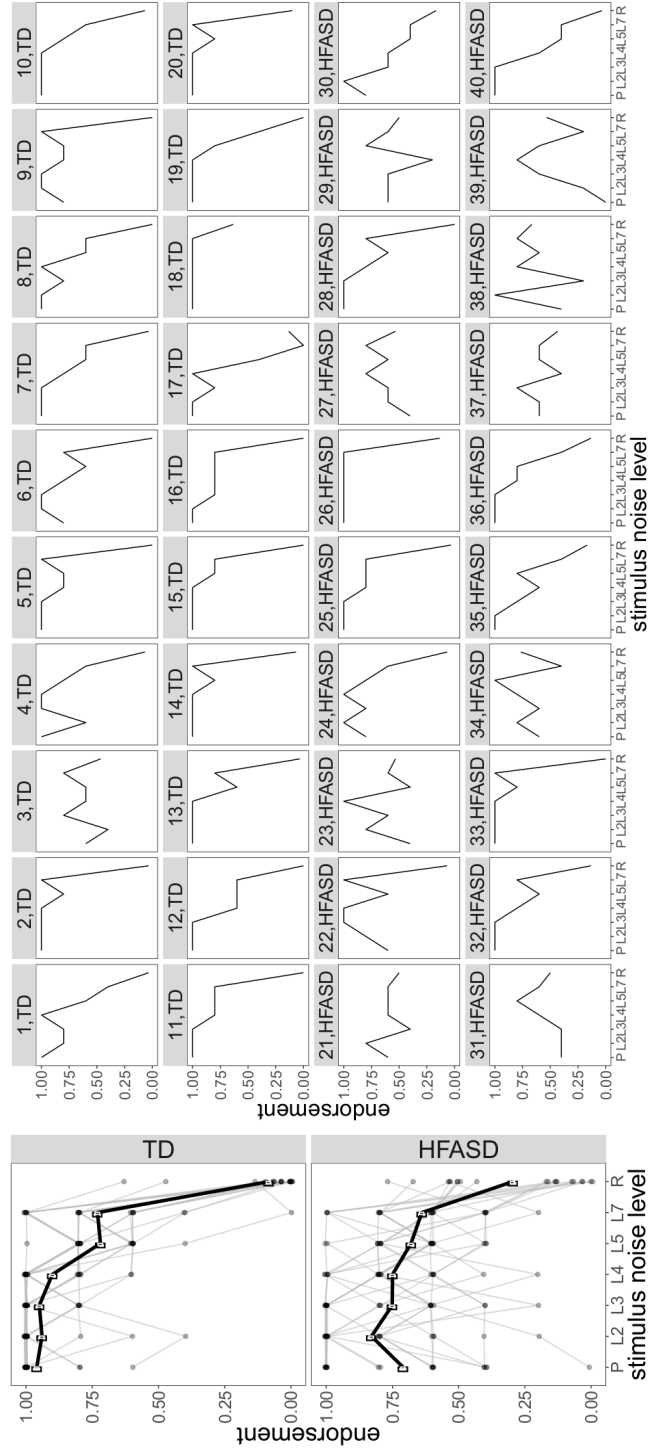


Figure 1: The left graph presents the average endorsement (solid black line) of stimuli varying in distortion level, as a function of group (TD: typically developing, HFASD: High-functioning Autism Spectrum Disorder). The right graph shows the average endorsement of the stimuli separately for each participant. Participants 1-20 were in the TD group, participants 21-40 in the HFASD group. The identification of participants is consistent throughout the manuscript to relate the different graphs.

29 and 34). These patterns provide reason to suspect that these children were using a different strategy when doing the task, which will be an important motivation for the second analysis.

### 3.2. *Categorization model*

Church et al. (2010) used a computational model of category learning, the *multiplicative prototype model* (MPM; Nosofsky, 1987; Smith & Minda, 1998), to understand the processes at play in the categorization task. According to the MPM, the probability of assigning a (novel) stimulus to a category depends on the similarity between the stimulus and the category's prototype, which is the average category member. Thus, when presented with members of a category, it is assumed that learners retain an abstract representation of that category by averaging across all the members. When then asked to judge whether a novel object is a category member, the learner evaluates the similarity of the novel object to the prototype and decides whether it is sufficiently similar.

Often, stimuli in a categorization task are represented as points in a similarity space, spanned by the physical stimulus dimensions. The psychological similarity between two stimuli is inversely related to the distance of the corresponding points in that space. The prototype is conveniently assumed to be the average of all category members in the similarity space. A crucial parameter in the prototype model reflects the sensitivity of participants to the distances in the similarity space when classifying the stimuli, that is, the precise shape of the function that maps physical differences (as represented in the similarity space) to psychological proximity that is used for categorization. The physical distances to the prototype in the similarity space

are given by the level of distortion from the prototype and can be found in Church et al. (2010, see appendix).

Formally, the psychological similarity  $\eta_j$  between the prototype and the  $j^{th}$  stimulus (denoted as  $S_j$ ) is assumed to be an exponentially decaying function of their physical distance  $d_j$  to the prototype:

$$\eta_j = e^{-cd_j}, \quad (1)$$

where  $c$  is a free parameter reflecting the sensitivity. The probability of endorsing a stimulus  $j$  as a category member is:

$$p(E_j = 1|S_j) = \frac{\eta_j}{\eta_j + e^{-q}} \quad (2)$$

where  $q$  is a free parameter reflecting the bias towards classifying a stimulus as a category member. The model used in Church et al. (2010) used  $p(E_j = 1|S_j) = \frac{\eta_j}{\eta_j + k}$ , with the free parameter  $k$  called the criterion. Our version is related to their version through re-parametrization as follows:  $k = e^{-q}$ .

### 3.2.1. Interpretation of the parameters

When the aim is to differentiate individuals and groups in terms of parameters in a model, it is crucial to have a clear and unambiguous interpretation of these parameters. In the prototype model, the parameter  $q$  can be interpreted as the log odds of classifying the prototype stimulus as a prototype. This follows from considering the prototype as a stimulus in Equation 2. In this case, the distance  $d_j$  is zero, the similarity  $\eta_j$  is one, and the probability to classify the prototype as a category member is:

$$p(E_j = 1|\text{prototype}) = \frac{1}{1 + \exp(-q)}. \quad (3)$$

Rewriting Equation 3 yields the desired interpretation of  $q$ :

$$q = \log \frac{p(E_j = 1|\text{prototype})}{1 - p(E_j = 1|\text{prototype})} = \log \frac{p(E_j = 1|\text{prototype})}{p(E_j = 0|\text{prototype})}. \quad (4)$$

If  $q = 0$ , the prototype is classified as a category member at chance level (see Figure 2). If  $q = \infty$ , this probability equals one, but even more moderate values of  $q$  yield very high classification probabilities. For example, a  $q$  of 3 implies a classification probability of 0.95. While  $q$  can in principle be negative, this is theoretically implausible as performance then drops below guessing.

Of more importance for the present purpose is the parameter  $c$ , which controls how physical and psychological distance are related. In particular, it reflects how sensitive participants are to the prototype. As illustrated in Figure 2, when  $c$  is large, only stimuli that lie very close to the prototype have non-negligible classification probabilities; when  $c$  is very small, all distortion levels yield the same classification probabilities, independent of their distances to the prototype.

#### 4. First Analysis: Hierarchical prototype model

Church et al. (2010) fit the prototype model to each individual’s data using SSD (sum of squared deviations), and then averaged the children’s parameter estimates within each group to study group differences. We recast the computational analysis of Church et al. (2010) in a hierarchical Bayesian framework, which considers all data simultaneously: On the basis of the categorization data, we infer the sensitivity to the prototype for each individual child separately, and, simultaneously, the group-level sensitivity to the prototype, separately for HFASD and TD children.

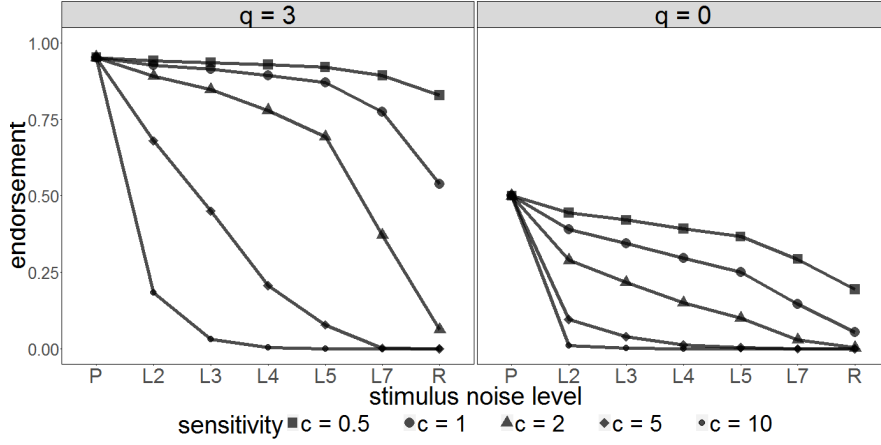


Figure 2: Classification probability as a function of log odds ( $q$ ) and sensitivity ( $c$ ). Log odds determines the endorsement probability at  $d = 0$ . Sensitivity influences the endorsement probability when distance to the prototype increases, with higher sensitivity resulting in lower endorsement.

By assuming separate population distributions for the HFASD and TD children, inferences are made separately for each group. However, we do not assume that all participants within a group have the same sensitivity, nor do we ignore similarities in sensitivity between participants within groups. Rather, we acknowledge that children within a group are different, without ignoring their similarities. The size of the differences between groups is based on the comparison of the population level distributions of the parameters of interest.

#### 4.1. Generative model

In the test phase of the categorization task, participants provided classification judgments for stimuli of varying similarity to the prototype. The data take the form of counts — the number of times a child  $k$  endorsed stimuli of a particular noise level  $j$  as a category member — which are assumed to

follow a binomial distribution:

$$y_{kj} \sim \text{Binomial}(r_{kj}, n_j) \quad (5)$$

where  $r_{kj}$  is the response probability and  $n_j$  is the number of stimuli of noise level  $j$  ( $j = 1, \dots, 7$ ). We assume the classification decisions follow a prototype process, which means that the response probability is calculated on the basis of the psychological distance  $d_j$  of stimulus  $j$  to the prototype such that

$$r_{kj} = \frac{e^{-c_k d_j}}{e^{-c_k d_j} + e^{-q_k}} \quad (6)$$

which follows directly from Equation 2.

The prototype model, with individual subject parameters  $c_k$  and  $q_k$  (with  $k = 1, \dots, 40$ ), is extended hierarchically with group-level parameters that describe the overall sensitivity and log odds of the HFASD and TD children:

$$(c_k, q_k) \sim \mathcal{N}(\mu^g, \Sigma^g) \quad (7)$$

characterized by a group-specific mean vector  $\mu^g = (\mu_c^g, \mu_q^g)$  and a by a group specific 2-by-2 covariance matrix  $\Sigma^g$ :

$$\Sigma^g = \begin{bmatrix} \sigma_c^{g2} & \sigma_{cq}^g \\ \sigma_{cq}^g & \sigma_q^{g2} \end{bmatrix}. \quad (8)$$

Depending on group membership (either HFASD or TD), the individual parameter is drawn from the respective group-specific bivariate normal distribution. Sensitivity and log odds capture separate aspects of the degree of learning a category, but can be correlated. The covariance matrix of the distribution is group specific, in the sense that separate variances and covariances are allowed for each group. In addition, the two groups may differ from one another in their mean vector:

$$\mu^g = \alpha + \delta G_k \quad (9)$$

where  $G_k$  indicates group membership of child  $k$ , taking on the value 1 if the child belongs to the HFASD group and 0 otherwise. The vector  $\alpha = (\alpha_c, \alpha_q)$  is the mean of the TD group (for a child for which  $G_k = 0$ ). The vector  $\delta = (\delta_c, \delta_q)$  reflects the difference in the mean of both parameters between the HFASD and TD group.

Broad priors are used for the means of the group level parameters:

$$\begin{aligned} \alpha_c &\sim \text{Uniform}(0, 10) \\ \alpha_q &\sim \text{Uniform}(0, 10) \\ \delta_c &\sim \mathcal{N}(0, 10) \\ \delta_q &\sim \mathcal{N}(0, 10) \end{aligned} \quad (10)$$

Group-level variance parameters  $\sigma_c$  and  $\sigma_q$  are given uniform priors between 0 and 10. We do not put a prior directly on the covariance. Instead, we specify a prior for the correlation parameter. This prior is a scale-location-transformed Beta(2,2) distribution, such that  $\frac{\rho+1}{2} \sim \text{Beta}(2, 2)$ .

The generative model exhaustively describes how classification data are produced. It is conceptually straightforward to reverse the generative process to infer which parameter values have likely produced the observed data, thus updating the prior distributions on the parameters to posteriors containing the updated knowledge about the parameters on the basis of the observed response patterns.

To check whether choice of prior had any notable influence on the posterior distributions, we conducted sensitivity analyses by using different con-

stants and distributions for the group-level variances. Whether uniform distributions (on  $\alpha_c$  and  $\alpha_q$ ) assumed an upper bound of 10, 20 or 50, results were essentially identical, as was the case for different specification of variance in the normal distributions of the difference between HFASD and TD children (variance of 10, 20, 50). Using gamma priors on the precision instead of uniform priors on the variances also yielded qualitatively identical results. Results are presented for the priors as specified above.

#### 4.2. Results

The model parameters were estimated in JAGS (Plummer, 2003). We ran 3 chains of 1,000,000 samples each, with a burn-in of 10,000 samples, withholding every tenth sample. Each chain was initialized with different values, drawn from the prior. Effective sample size was larger than 22,000 for all parameters. Convergence was established visually for the group level parameters and by calculating  $\hat{R}$ . For all parameters,  $\hat{R}$ , which is an evaluation of the stability of parameter estimates across different chains on the basis of within chain and between chain variability (Brooks & Gelman, 1997) was smaller than 1.001.

Figure 3 shows the joint posterior means for  $c_k$  and  $q_k$  at the participants' level (left graph) and the participants' posterior means of the main parameter of interest, the sensitivity parameter  $c_k$  with the 95% highest density interval (right graph), broken up by group membership. Clearly, there is a strong correlation between sensitivity and log odds across participants. Further, two clusters of participants emerge: One cluster includes participants with very low estimated values of both sensitivity and log odds, while a second cluster comprises participants with higher estimates on these parameters.



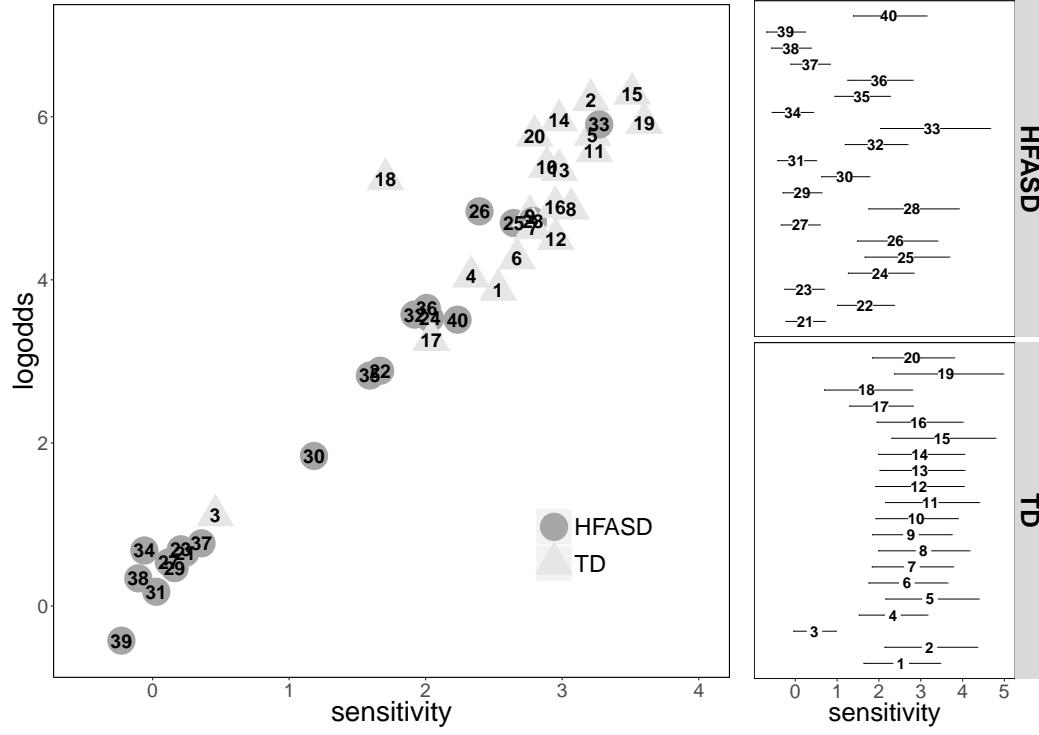


Figure 3: Joint posterior subject means for sensitivity  $c$  and logodds  $q$  for the two groups (left) and posterior means and 95% highest density intervals for  $c$  (right).

All but one TD child (participant 3) are in the high estimates cluster. The HFASD participants, in contrast, are about equally spread out across both clusters. The picture of a relatively homogenous set of estimates for the TD group (with the exception of participant 3) and at least two groups of estimates in the HFASD group is confirmed when zooming in on sensitivity (Figure 3, right graph).

The left graph in Figure 4 presents the joint and marginal posterior distribution of the group-level means. The marginal posterior distributions reveal a substantial difference between groups for both parameters. HFASD chil-

dren appear to be lower in both log odds of classifying the prototype as a category member, as well as in sensitivity to differences in the stimulus space. Zooming in on sensitivity, the posterior distribution on  $\delta_c$ , which quantifies the difference between TD children and HFASD children on the sensitivity parameter, supports this conclusion. The mean difference is estimated to be  $-1.47$ , with a 95% highest density interval of  $[-2.26; -0.68]$ .

A formal test of the difference can be performed by calculating a Bayes factor that compares two models: a model that allows a difference between TD and HFASD children in the sensitivity parameter ( $\delta_c \neq 0$ ), and a model that assumes there is no difference ( $\delta_c = 0$ ). Applying the Savage-Dickey procedure (Wagenmakers et al., 2010) to  $\delta_c$  we found a Bayes factor  $> 70$  in favor of the difference model. A sensitivity analysis revealed that the Bayes factor decreased when broadening the priors, but remained in favor of a difference even in the broadest settings ( $BF = 28$  when  $\delta_c \sim \mathcal{N}(0, 50)$ ).

#### *4.3. Discussion*

Our first analysis is very similar to the analysis by Church et al. (2010). Like our analysis, their analysis revealed lower sensitivity parameter estimates for the HFASD group (a difference of 0.97, see their Table 2). HFASD children seem to resolve distances from the prototype less sensitively than TD children.

An important difference with the analysis of Church et al. (2010) is the Bayesian hierarchical approach taken here. Our analysis produces a posterior distribution of the difference between the HFASD and TD population means for the sensitivity and log odds parameter. Instead of only having individuals' parameters, or only parameter estimates on the basis of averaged data

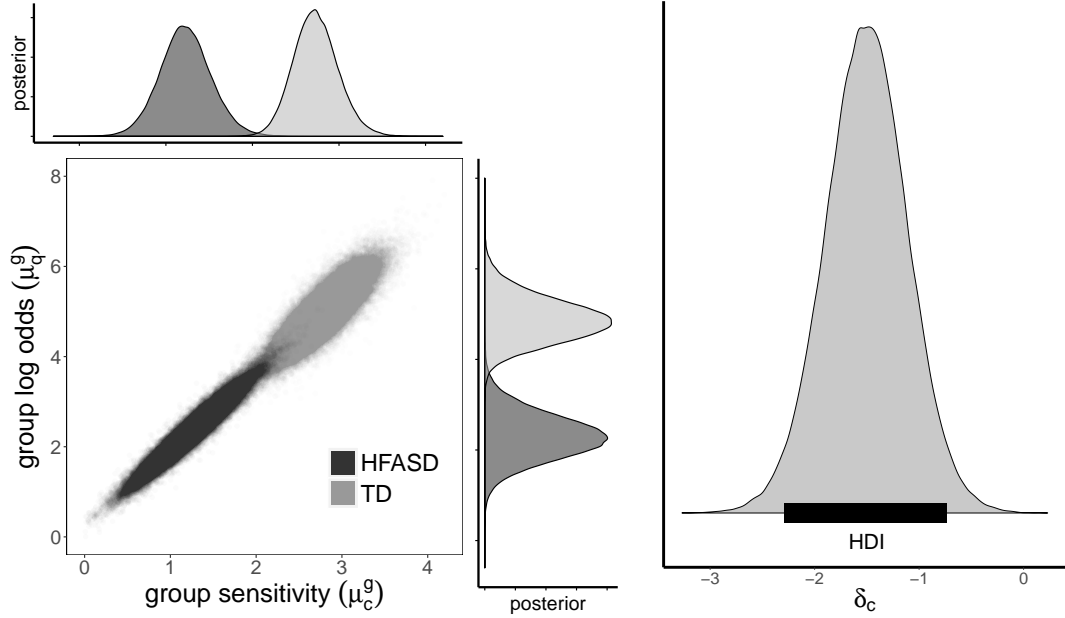


Figure 4: The left panel shows the joint and marginal posteriors of the group level mean parameters for TD and HFASD children. The right graph shows the posterior on  $\delta_c$ , indicating the difference between the group level sensitivity parameter for HFASD and TD, and the 95% highest density interval.

per group, the Bayesian hierarchical analysis simultaneously yields both, allowing differences between children within a group (see Figure 3) as well as acknowledging the commonality of children of in a particular group (see Figure 4). Both levels inform each other in the inference process.

Before interpreting parameter estimates, it is necessary to make sure the model is appropriate for the data. In our case, there are reasons to suspect that the prototype model is not appropriate for all participants. In particular, we have already highlighted that some response profiles in Figure 1 (e.g., participants 3 and 37) do not reflect the generalization gradient that most children display, and that one expects to see (that is, generally decreasing endorsement as stimulus noise levels increase). These children appear to be

responding at chance level, guessing independent of stimulus noise level. As shown in Figure 3, they also form the clusters with low estimates for both parameter  $c$  and  $q$ .

In sum, there appear to be (at least) two subgroups of HFASD children and one subgroup may not rely on prototype processes, with their responses hovering around chance level. Importantly, the validity of any parameter inference requires the model to be valid to a reasonable extent: The data of participants who did not follow a prototype strategy, should not be used when estimating prototype model parameters.

## 5. Second Analysis: Hierarchical mixture prototype model

If a number of participants were not relying on a prototype strategy, but guessed on each trial, the population level parameter estimates are a mixture of parameter values of prototype classifiers and guessers. To isolate those participants for which the prototype model is not appropriate, we will compare the prototype model, as specified by Equations 1 and 2, to a simple guessing model, which assumes a probability of endorsement of 0.5 for each stimulus type, irrespective of its relation to the prototype. If the data of a participant are better captured by the guessing model than by the prototype model, we take this as evidence that the prototype model is not appropriate for that participant, and that the estimates of the sensitivity parameter are useless for the participant.

In particular, we assume two clusters: a prototype cluster comprising children that relied on a prototype process and a guessing cluster comprising children that appear to be guessing (Bartlema et al., 2014; Zeigenfuse

& Lee, 2010). The cluster assignments are latent, which means that participants are assigned to a cluster by the model and not on the basis of some observed value. As we are interested in differences between TD and HFASD children in the prototype parameters, we assume there are two subgroups in the prototype cluster: TD and HFASD children.

### 5.1. Generative model

Formally, we again assume that the data follow a binomial distribution with an endorsement probability  $r_{kj}$ , of participant  $k$  of a stimulus of noise level  $j$ . However, different from the previous analysis,  $r_{kj}$  either depends on a guessing strategy or on a prototype model process:

$$r_{kj} = z_k \frac{1}{2} + (1 - z_k) \left( \frac{e^{-c_k d_j}}{e^{-c_k d_j} + e^{-q_k}} \right) \quad (11)$$

with  $z_k$  indicating whether participant  $k$  follows a guessing strategy or a prototype strategy.

The probability with which a participant is relying on a guessing strategy is controlled by a Bernoulli process with a separate, group-specific rate  $\theta^g$ , with  $g$  indicating group membership (TD or HFASD):

$$z_k \sim \text{Bernoulli}(\theta^g) \quad (12)$$

with a noncommittal prior on the parameters  $\theta^g \sim \text{Beta}(1, 1)$ . If a participant is assigned to the guessing cluster ( $z_k = 1$ ), all classification judgments are made with a response rate of .5. If a participant is assigned to the prototype cluster ( $z_k = 0$ ), classification judgments follow the prototype model described earlier. Reversing the flow of the generative model, we can infer whether a participant followed a guessing or prototype strategy (the posterior

probability of  $z_k$ ) on the basis of her response pattern, simultaneously establishing the probability ( $\theta^g$ ) that HFASD children and TD children resorted to guessing.

To check whether choice of prior had any notable influence on the resulting posterior distribution, priors were varied as in the first analysis, yielding qualitatively identical results. Results are presented for the priors above.

## 5.2. Results

We ran 3 chains of 1,000,000 samples each, with a burn-in of 10,000 samples, keeping every tenth sample. Each chain was initialized with different parameter values, randomly drawn from the priors. Effective sample size was larger than 25,000 for all parameters. Convergence was established visually for the group level parameters and the  $\hat{R}$  measure for all parameters was smaller than 1.01.

Figure 5 shows the latent assignment of participants, indicated by the mean of a binary latent group indicator ( $z_k$ ), and the posterior of being assigned to the guessing strategy ( $\theta^g$ ). The right panel shows that the probability of resorting to a guessing strategy is larger for HFASD children (mean posterior probability of .41, with 95% highest density interval [0; 0.21]) than for TD children (.09, [0.20; 0.63]).

At the individual level, for all but one participant (participant 34 from the HFASD group), assignments are very outspoken. Eight children with HFASD and one TD child (participant 3) are better accounted for by a simple guessing model than by the prototype model. For these children, the prototype model is clearly not the most appropriate model. This observation has severe implications for the research question on sensitivity differences between

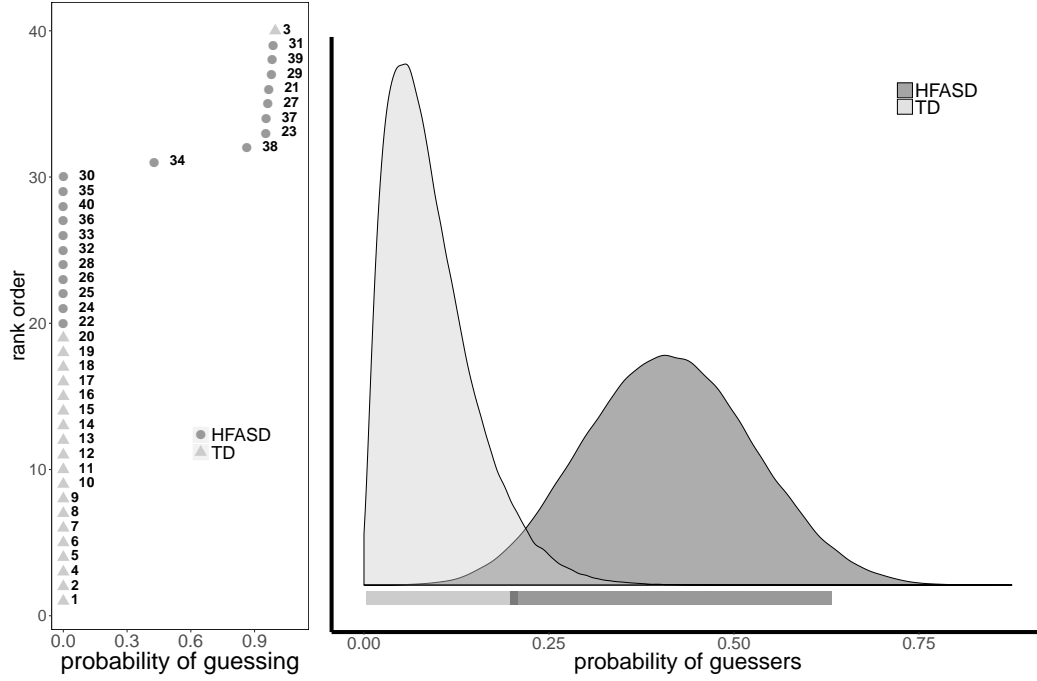


Figure 5: Latent assignment to the guessing group (left graph). The right graph shows the posterior on  $\theta$  (probability of being assigned to the guessing strategy) for each group. The bars under the graph represent the 95% highest density interval for both groups.

HFASD and TD children: The data of the children for which the prototype model is not appropriate should not inform the population parameter estimates of sensitivity and log odds. Results are more uncertain for participant 34, who is included in the prototype group in approximately 40% of the posterior samples, thus influencing inference on higher level distributions only in those samples. In the following graphs, we assign participant 34 to the guessing group, as the posterior probability indicates it is uncertain whether he or she was following a prototype strategy.

Focusing on the participants for which the prototype model was appropriate, or at least more appropriate than a simple guessing model, Figure 6

presents the joint posterior means for sensitivity and log odds. The children with extremely low  $c_k$  and  $q_k$  values in Figure 3 are no longer depicted, as they are now assigned to the guessing cluster. Apparently, our simple guessing model was able to capture their data profiles better than the prototype model. These participants will no longer influence the population level parameter estimates (or only marginally so, depending on the posterior probability of being assigned to the guessing group).

Taking a further look at group level posterior distributions (Figure 7), it is clear that the marginal distributions of the population means ( $\mu_c^g$  and  $\mu_q^g$ ) overlap more than in the first analyses. In other words, HFASD and TD children are not as easily differentiated with respect to sensitivity and log odds. Zooming in on sensitivity, the posterior mean for the difference between the two groups is  $-.68$ , with the 95% highest density interval being  $[-1.49; 0.07]$ . Comparison of a model without difference in sensitivity between TD and HFASD and the model with — relying on the Savage-Dickey procedure on the basis of  $\delta_c$  — yielded a Bayes Factor of 1.6 in favor of the model assuming there is no difference (i.e.,  $\delta_c = 0$ ). As expected, the Bayes factor slightly increased with broader priors, to 3.6 when  $\delta_c \sim \mathcal{N}(0, 50)$ . While providing weak support for the absence of a difference in sensitivity between TD and HFASD, the Bayes factor mainly indicates that the data are not sufficiently informative to decide between the two models.

### 5.3. Discussion

To get a sense of the extent to which the model can account for the data, Figure 8 presents the group level and individual level posterior predictive, together with the observed data. A posterior predictive check is a Bayesian



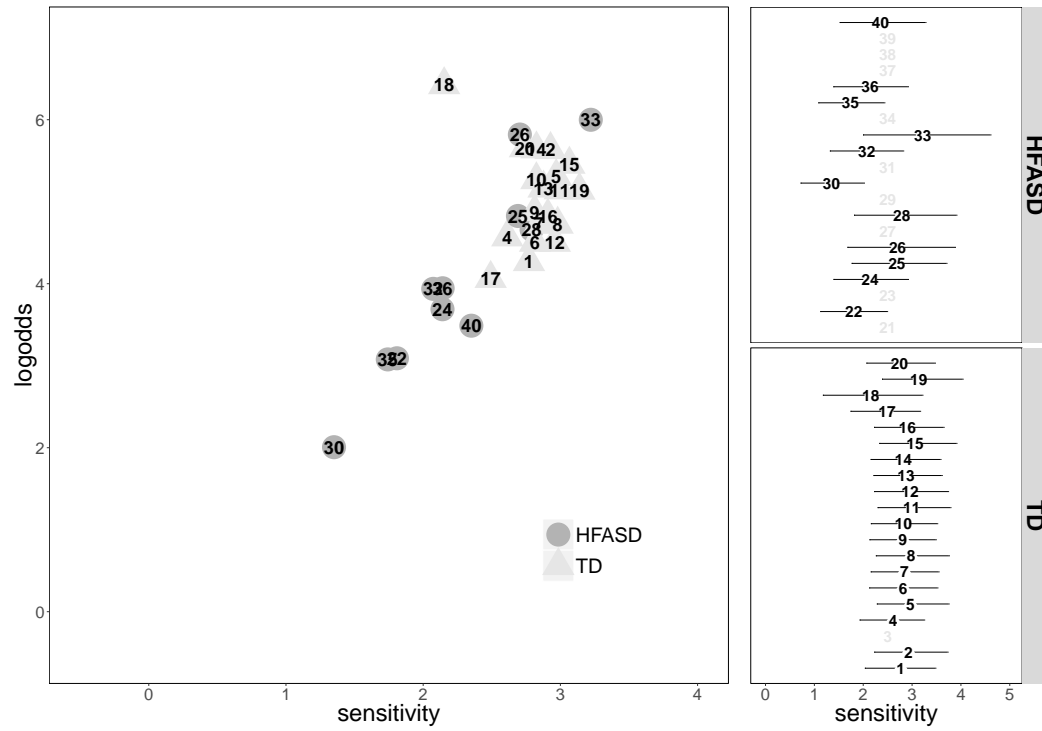


Figure 6: Joint posterior subject means for sensitivity  $c$  and logodds  $q$  for the two groups (left) and posterior means and 95% highest density intervals for sensitivity. The graphs only shows participants whose assignment to the prototype strategy was clear.

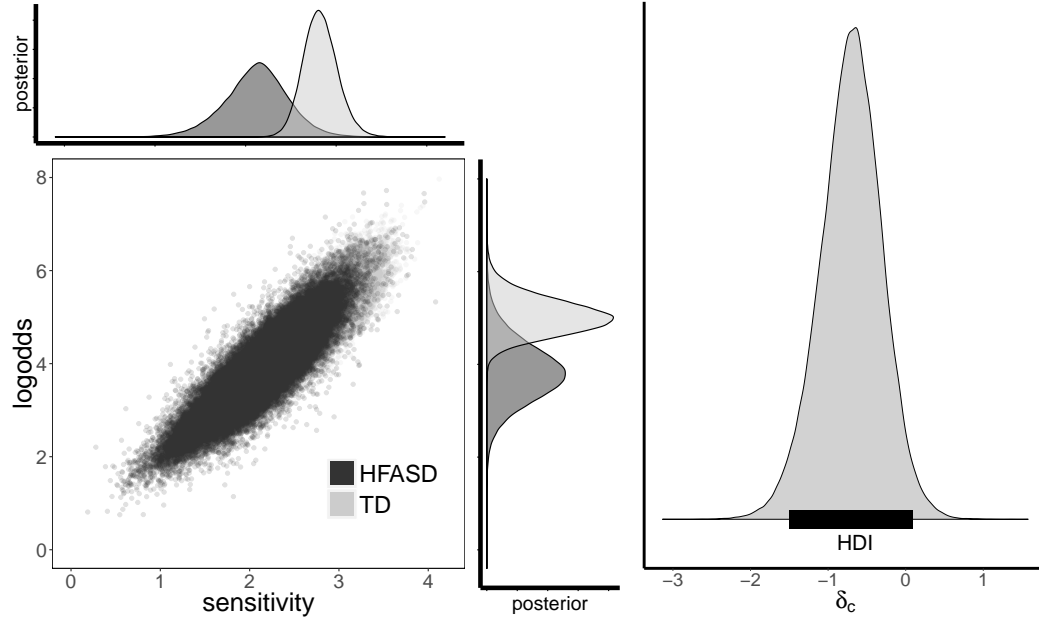


Figure 7: Joint and marginal posterior distribution of the mean of the group distribution for the TD subjects and HFASD subjects (left). The right graph presents the posterior on the difference between HFASD and TD subjects in terms of the sensitivity parameter (at a group level), the black bar indicating the 95% credible interval.

method for assessing the fit of a model to data, based on simulating data from the posterior predictive distribution and comparing these simulations to the observed data (Gelman & Hill, 2006; Lee & Wagenmakers, 2014; Shiffrin et al., 2008). As can be seen from the left panel, the guessing group comprises the participants that were most erratic in their categorization responses, reflecting little evidence for the generalization gradient one would expect in a classification task. The data patterns of the prototype groups, TD and HFASD children, seem more homogeneous, and do reflect the generalization gradient. Not surprisingly, the differences in the parameter posteriors have decreased. By isolating the guessers, the response patterns are also consistent

with the analysis' assumption of homogeneous groups.

The guessing model is extremely simple, without any free parameters, which gives it an advantage in model comparisons with the prototype model that has more flexibility with two free parameters (Myung, 2000). Yet, response patterns that reflect the general decrease in endorsement as a function of similarity are unambiguously assigned to the prototype model. Thus, despite its higher complexity, the prototype model compares favorably to the simple guessing model in these cases. Also, latent assignments of children were robust against different prior settings in the sensitivity analysis, even with the most diffuse priors for the prototype model.

Our analysis reveals two types within the HFASD group: One group whose behavior can be parsimoniously captured by assuming they are simply guessing, and one group whose behavior relies on abstraction of a prototype when learning about the category, and their classification judgments derive from similarity to the abstracted prototype. For the children assigned to the guessing group, the prototype model is not appropriate and, consequently, parameter estimation using that model does not make sense. This has the epistemological implication that for these children, we are unable to assess the sensitivity to the prototype. For the children assigned to the prototype group, assessing the sensitivity to the prototype *is* possible. On the basis of their data alone, the difference in sensitivity between TD and HFASD children is not as outspoken. At the very least, care should be taken when inferring anything substantial regarding differences between HFASD and TD children in sensitivity to the prototype when classifying.

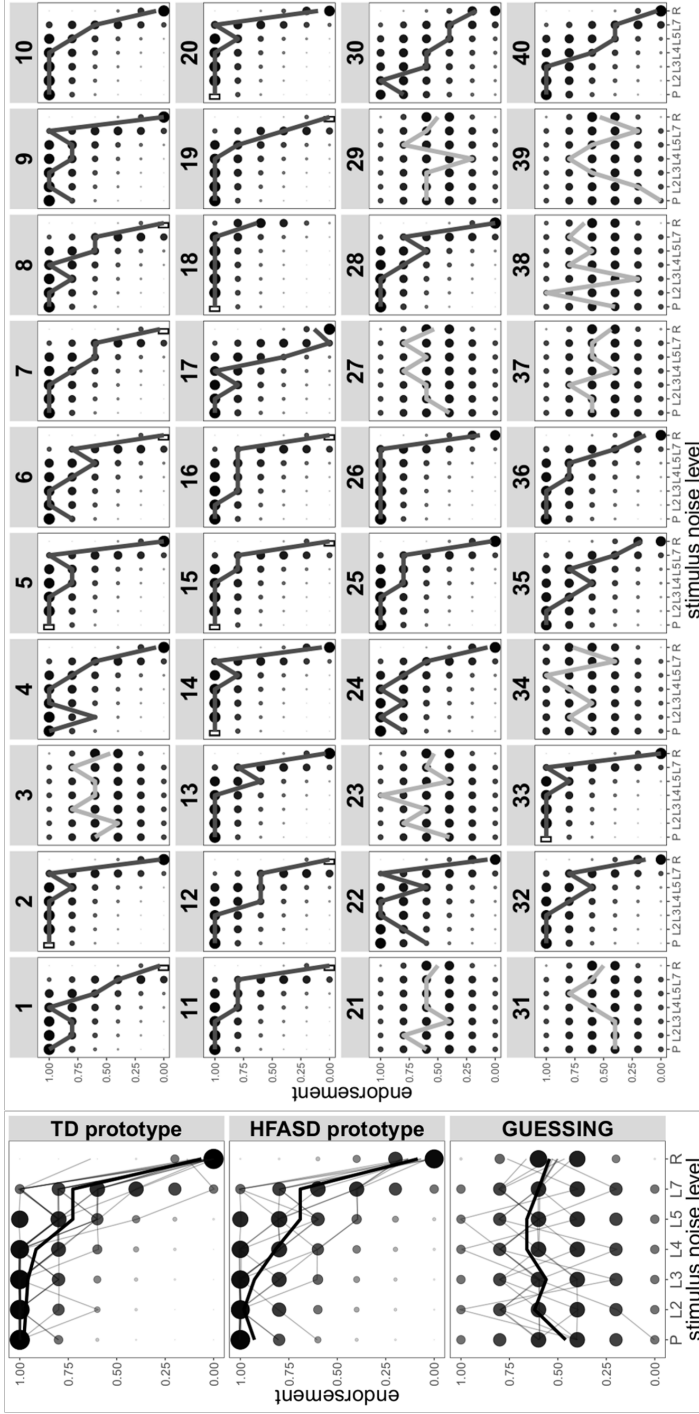


Figure 8: The left panel shows the posterior predictive at the group level (left panel), presenting the three groups: TD children, HFASD children and children that were not using a prototype representation (GUESSING). Solid black lines represent the group average proportion of endorsement, and the grey lines are individual data profiles for that group. The right panel shows the posterior predictive using individual level parameters. Children assigned to the guessing group have light grey data lines in the individual plots, the other children have black lines. For both left and right panel: The size and grey-value of the circles reflect the posterior probability of an observation, smaller and more grey values indicating lower probability.

## 6. General Discussion

In this paper, we presented a case study of hierarchical Bayesian explanatory cognitive psychometrics. We demonstrated how well-studied lab paradigms from cognitive science, in tandem with cognitive models and Bayesian hierarchical mixture methods, can be used to assess individual differences and relate them to covariates such as ASD. Our analysis combined different elements from a traditional IRT analysis (see e.g., De Boeck & Wilson, 2004). A statistical model was used to describe the response process. Individual differences were modeled using a hierarchical population distribution. Manifest (i.e., group assignment: HFASD vs. TD) and latent covariates (guessing vs. non-guessing) were used to explain the individual differences. The major difference is that the response process model was not a generic IRT model, but rather a computational model deeply rooted in cognitive theory: the prototype model.

The parameters in the prototype model reflect information processing characteristics that are theoretically related to ASD. To evaluate whether these characteristics are different for HFASD children, we assumed that they were drawn two different population distributions: one for HFASD children and one for TD children. In a first analysis, we found a difference in sensitivity to the prototype, with HFASD children being less sensitive. However, in a second analysis, we added a mixture component to isolate children that did not seem to rely on prototype processes. When focusing only on participants for which a prototype model seemed appropriate, the difference in sensitivity between HFASD and TD children largely disappeared, which is in line with recent evidence suggesting that HFASD children that extract a prototype are

near-identical in sensitivity to TD children (Church et al., 2015). Yet, the Bayes factor testing for a difference primarily indicated that the present data are not sufficiently informative to decide either way.

So why did we find evidence in favor of a difference in our first analysis? This can for a large part be attributed to the presence of a subgroup of participants who seemed to have been guessing. The prototype model can capture the response profiles of these participants to a certain extent, by setting the sensitivity and log odds parameters close to zero. With near-zero sensitivity and log odds, response probabilities are close to .5, independent of the similarity of the stimulus to the prototype. Because the guessers were not isolated in the first analysis, and the guessing strategy was substantially more frequent in the HFASD group, their group-level sensitivity and log odds were substantially lowered. In the second analysis, we identified the guessers, and they were not used to infer the inference about the prototype model’s parameters. As a result, the difference in sensitivity was no longer convincing.

It may be disappointing that we are not able to provide more definite answers to the questions formulated by Church et al. (2010). However, the inferences made here are the best we can possibly make at this point. Stronger conclusions would be misleading. We consider it a strength of the Bayesian hierarchical mixture approach that it heeds us from overly strong conclusions. That is not to say that stronger inferences are impossible: Increasing the strength of either the theory side — by making stronger assumptions regarding processes, subgroups, etc... — or the empirical side — by gathering more data, or by, for example, building a more elaborate stimulus space —

might lead to stronger conclusions.

### *6.1. Interpretation of the guessing group*

The existence of different subgroups within the HFASD population should not come as a surprise (see, e.g., Caron et al., 2006; Church et al., 2015; Rajendran & Mitchell, 2007) and corroborates earlier evidence of multiple response patterns in these data (Dovgopoly & Mercado III, 2013). In our second analysis, about half of the HFASD children were assigned to a guessing group rather than the prototype group. This is not to say that we are sure that these children were in fact guessing. It simply means that a simple model with no free parameters, setting all response probabilities at 0.5, can accommodate the data at least equally well as the substantially more complex prototype model. In this light, it is unwise to use the prototype model to understand their behavior.

It has been argued that social and non-social symptoms that are generally associated with ASD are not necessarily inherently linked and may have very different causes at the genetic, neural and cognitive level (the fractionable autism triad hypothesis, see, e.g., Brunsdon & Happé, 2014; Happé & Ronald, 2008). Even within the cognitive domain, symptoms related to executive functioning on the one hand and focus on detail and the difficulty to integrate information on the other may not be caused by the same underlying principles (see e.g., Best et al., 2008; Happé & Booth, 2008; Lawson et al., 2004). The present findings are consistent with this hypothesis, as different subgroups were clearly identified in the classification behavior. Unfortunately the interpretation of the subgroups is not straightforward, as the guessing patterns can have a variety of causes, ranging from impaired as-

signment of cognitive resources (an executive function) to prototype related processes such as the extent to which a sensible prototype is abstracted in the learning phase, and the extent to which the distances in the psychological space are stretched or shrunk as controlled by the sensitivity parameter.

### *6.2. Sensitivity and hyperspecificity*

A crucial step in cognitive psychometrics is determining what the parameter estimates mean from a theoretical perspective. Unfortunately, the relationship between the sensitivity parameter, as build in the prototype model, and current theories on ASD is not as clear as one would hope. From the perspective of enhanced perceptual functioning of individuals with ASD (e.g., Mottron et al., 2006) one expects the formation of highly specific and detailed representations. The exact modelling translation of hyperspecificity, however, is unclear. One translation implies the formation of exemplar based representations rather than prototype representations (Medin & Schaffer, 1978). In this scenario, one expects a lower sensitivity to the prototype for HFASD than for TD controls — one can not be sensitive to what has not been formed (although as mentioned above, applying the prototype model when assuming no prototype has been formed is highly questionable). This seems to be the reasoning used in Church et al. (2010). Another translation could be that prototypes can be formed. In that case, one expects larger sensitivity values for the HFASD participants: Larger sensitivity values stretch the stimulus space, exaggerating differences, leading to the detailed, piecemeal representations following from enhanced perceptual functioning (see Bott et al., 2006).

Clearly, the application of the computational prototype model forces re-



searchers, and rightfully so, to be very explicit regarding which aspect is expected to be different in individuals with ASD. In light of this, it should be a key concern in further studies of categorization performance of individuals with ASD, to develop stimuli, category structures and designs that allow differentiation of the prototype formation and sensitivity, given their apparent importance in theories regarding ASD.

### *6.3. Conclusion*

We have illustrated how Bayesian hierarchical mixture models can be usefully applied in explanatory cognitive psychometrics. The goal in this approach is to use cognitive models to produce estimates of psychologically interesting parameters regarding cognitive functioning, and to relate them to external co-variates. By re-analysis of data collected by Church et al. (2010) on categorization in children with HFASD, we have demonstrated that this approach can lead to nuanced answers to the questions at hand and provides information on uncertainty regarding these answers. Throughout, we have highlighted the importance of establishing the appropriateness of the model before taking the parameter estimates seriously. The resulting answers can be considered the best possible answers, given the question at hand, the level of theoretical development and the data collected.

## **7. Acknowledgements**

We thank Barbara Church for generously providing us with the categorization data and fruitful input. Wouter Voorspoels is a postdoctoral researcher at the KU Leuven. Isa Rutten is a graduate student at KU Leuven. The research leading to the results reported in this paper was supported

in part by the Research Fund of KU Leuven (GOA/15/003) and by the Interuniversity Attraction Poles programme financed by the Belgian government (IAP/P7/06).

## 8. References

- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. Washington, DC: Author.
- Bartlema, A., Lee, M. D., Wetzels, R., & Vanpaemel, W. (2014). A Bayesian hierarchical mixture approach to individual differences: Case studies in selective attention and representation in category learning. *Journal of Mathematical Psychology*, *59*, 132–150.
- Batchelder, W. H. (1998). Multinomial processing tree models and psychological assessment. *Psychological Assessment*, *10*, 331–344.
- Best, C., Moffat, V., Power, M., Owens, D. G. C., & Johnstone, E. C. (2008). The boundarie of the cognitive phenotype of autism: Theory of mind, central coherence and ambiguous figure perception in young people with autistic traits. *Journal of Autism and Developmental Disorders*, *38*, 840–847.
- Bott, L., Brock, J., Brockdorff, N., Boucher, J., & Lamberts, K. (2006). Perceptual similarity in autism. *Quarterly Journal of Experimental Psychology*, *59*, 1237–1254.
- Brooks, S. P., & Gelman, A. (1997). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, (pp. 434–455).

- Brunsdon, V. A., & Happé, F. F. (2014). Exploring the 'fractionation' of autism at the cognitive level. *Autism*, 18, 17–30.
- Caron, M., Mottron, L., & Dawson, M. (2006). Cognitive mechanisms, specificity and neural underpinning of visuospatial peaks in autism. *Brain*, 129, 1789–1802.
- Church, B. A., Krauss, M. S., Lopata, C., Toomey, J. A., Thomeer, M. L., Coutinho, M. V., Volker, M. A., & Mercado III, E. (2010). Atypical categorization in children with high-functioning autism spectrum disorder. *Psychonomic Bulletin & Review*, 17, 862–868.
- Church, B. A., Rice, C., Dovgopoly, A., Lopata, C., Thomeer, M. L., Nelson, A., & Mercado III, E. (2015). Learning, plasticity, and atypical generalization in children with autism. *Psychonomic Bulletin & Review*, (pp. 1–7).
- De Boeck, P., & Wilson, M. (2004). *A framework for item response models: A generalized linear and non-linear approach..* Springer.
- Dovgopoly, A., & Mercado III, E. (2013). A connectionist model of category learning by individuals with high-functioning autism spectrum disorder. *Cognitive, Affective, & Behavioral Neuroscience*, 13, 371–389.
- Gelman, A., & Hill, J. (2006). *Data analysis using regression and multi-level/hierarchical models*. NY: Cambridge University Press.
- Happé, F., & Booth, R. (2008). The power of the positive: Revisiting weak coherence in autism spectrum disorder. *Quarterly Journal of Experimental Psychology*, 61, 50–63.

- Happé, F., F., & Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 36, 5–25.
- Happé, F. F., & Ronald, A. (2008). The 'fractionable autism triad': A review of evidence from behavioural, genetic, cognitive and neural research. *Neuropsychology Review*, 18, 287–304.
- Knowlton, B., & Squire, L. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science*, 262, 1747–1749.
- Kruschke, J., & Vanpaemel, W. (2015). Bayesian estimation in hierarchical models. In J. Busemeyer, J. Townsend, Z. Wang, & A. Eidels (Eds.), *Oxford Handbook of Computational and Mathematical Psychology* Oxford Library of Psychology. Oxford: Oxford University Press.
- Lawson, J., Baron-Cohen, S., & Wheelwright, S. (2004). Empathising and systematising in adults with and without asperger syndrome. *Journal of Autism and Developmental Disorders*, 34, 301–310.
- Lee, M. (2008). Three case studies in the Bayesian analysis of cognitive models. *Psychonomic Bulletin & Review*, 15, 1–15.
- Lee, M. (2011). How cognitive science can benefit from hierarchical Bayesian models. *Journal of Mathematical Psychology*, 55, 1–7.
- Lee, M., & Vanpaemel, W. (submitted). Determining informative priors for cognitive models, .

- Lee, M. D. (in press). Bayesian methods in cognitive modeling. In *The Steven's Handbook of Experimental Psychology and Cognitive Neuroscience*. Wiley. (4th ed.).
- Lee, M. D., & Wagenmakers, E. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley Publishing Company.
- van der Maas, H., Molenaar, D., Maris, G., Kievit, R., & Borsboom, D. (2011). Cognitive psychology meets psychometric theory: on the relation between process models for decision making and latent variable models for individual differences. *Psychological Review*, *118*, 339–356.
- Medin, D., & Schaffer, M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207–238.
- Mottron, L., Dawson, M., Soulires, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual function in autism: An update and eight principles of autistic perception. *Journal of Autism and Developmental Disorders*, *36*, 27–43.
- Myung, I. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology*, *44*, 190–204.
- Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *13*, 87–108.

- Plaisted, K. C. (2001). Reduced generalization in autism: An alternative to weak central coherence. In K. C. Plaisted, J. Burack, T. Charman, N. Yirmiya, & P. Zelazo (Eds.), *The development of autism: Perspectives from theory and research* (pp. 149–169). Lawrence Erlbaum Associates Publishers.
- Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using gibbs sampling. In K. Hornik, F. Leisch, & A. Zeileis (Eds.), *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*. Vienna, Austria.
- Posner, M., & Keele, S. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, 77, 353–363.
- Rajendran, G., & Mitchell, P. (2007). Cognitive theories of autism. *Developmental Review*, 27, 224–260.
- Riefer, D. M., Knapp, B., Batchelder, W. H., Bamber, D., & Manifold, V. (2002). Cognitive ppsychometric: Assessing storage and retrieval deficits in special populations with multinomial processing tree models. *Psychological*, 14, 184–201.
- Rouder, J., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, 12, 573–604.
- Shiffrin, R. M., Lee, M., Kim, W., & Wagenmakers, E. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science*, 32, 1248–1284.

- Smith, J., & Minda, J. (1998). Prototypes in the mist: the early epochs of category learning. *Journal of Experimental Psychology*, *24*, 1411–1436.
- Tuerlinckx, F., & De Boeck, P. (2005). Two interpretations of the discrimination parameter. *Psychometrika*, *70*, 629–650.
- Vanpaemel, W., & Lee, M. D. (2012). Using priors to formal theory: Optimal attention and the Generalized Context Model. *Psychonomic Bulletin & Review*, *19*, 1047–1056.
- Verbeke, G., & Molenberghs, G. (2009). *Linear mixed models for longitudinal data*. Springer Verlag New York.
- Wagenmakers, E., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the savage-dickey method. *Cognitive Psychology*, *60*, 158–189.
- Zeigenfuse, M. D., & Lee, M. D. (2010). A general latent assignment approach for modeling psychological contaminants. *Journal of Mathematical Psychology*, *54*, 352–362.